The perception of two linguistic functions of prosody in Danish

Tøndering, John; Morris, David Jackson

Publication date:
2015

THE PERCEPTION OF TWO LINGUISTIC FUNCTIONS OF PROSODY IN DANISH

John Tøndering and David Morris

Department of Nordic Studies and Linguistics, University of Copenhagen
{johnt, dmorris@hum.ku.dk}

ABSTRACT

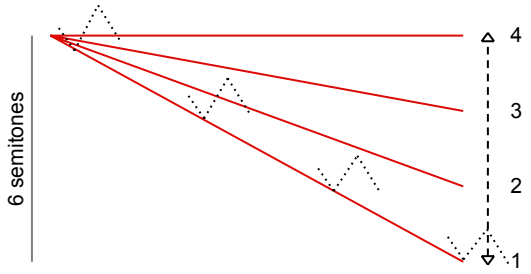
This paper examines the prosodic perceptual ability of listeners (n=22) in tasks involving global intonation contours and local voice pitch variation. The signaling of utterance type was tested with speech material that included read and non-scripted questions and statements. Local variation was tested with an identification task where the level of prominence of an anaphor signaled either the subject or non-subject of a preceding sentence. Logistic regression revealed that question and statement identification differed between read and non-scripted material and was not linked to performance on the subject/non-subject task. A response bias was also observed in the vocoded condition towards statements. We conclude that performance differences between the two tasks may be due either to the task-related language ability of the listener or to differences in the perception of global and local prosody.

Keywords: prosody, intonation, question, statement, vocoder.

1. INTRODUCTION

The current hierarchical model of intonation in read aloud Danish involves the superposition of shorter voice fundamental frequency (F_0) variations onto intonation contours that occur over a longer duration [4]. The longer intonation contours carry and scale the shorter F_0 variations. Thus, the stress group pattern, i.e., the F_0 pattern associated with stressed and subsequent unstressed syllables, is superposed on the utterance intonation contour. Utterances with more than 4-5 stressed syllables are divided into prosodic phrases. The prosodic phrase contours are overlaid on and constrained by the global utterance contour. The slope of the utterance contour signals utterance function according to Fig. 1. Terminal declarative sentences typically have a steeply falling utterance contour, and echo-questions a horizontal contour. Other questions and non-final clauses have a slope that lies between these. [2] reported that this intonation contour continuum can also be observed in non-scripted Danish, although questions with declarative word order had steeper descending contours than questions with word order inversion.

Figure 1: Intonation contour slopes (red lines) in: (1) declaratives; (2) wh-questions; (3) questions with word order inversion, and; (4) declarative questions.



In standard Copenhagen Danish, the stress group has a uniform F_0 pattern that is represented by the dotted lines in Fig. 1. It is characterized by a brief fall, then a rise to the first stressed syllable, followed by a fall through any succeeding unstressed syllables. The local minimum at the beginning of this stress group pattern coincides with the stressed vowel, and is attached to the global utterance contour. The prominence level of a stressed syllable, and of the word in which it occurs, is associated with a rise in F_0. The magnitude of this rise is a primary cue to the perceptual prominence of the syllable.

In the present study we examined whether listeners' ability to perceive a globally signaled function was related to their ability to derive meaning from a local prosodic cue. To assess this we investigated whether the ability to identify questions and statements, was related to performance on a task where the prominence level of a word modified the antecedent of a pronoun over a sentence boundary. Our hypothesis was that a listeners' sensitivity to the distinctive intonation contours of questions and statements would be related to their linguistic perceptual use of prominence. The ability to perceive the utterance type from the global contour was tested on the basis of both read and non-scripted speech. In testing these two modes of speech we were interested in the difference in the bias of listeners towards the statement answer choice, as the upper and lower

boundaries of the intonation contour continuum has been described as highly marked and unmarked, respectively [5, p. 26]. In order to observe the role of F_0 in these tasks we included two noiseband vocoder conditions: a full-spectrum vocoder and one that retained F_0 but marred other spectral detail via noiseband vocoding.

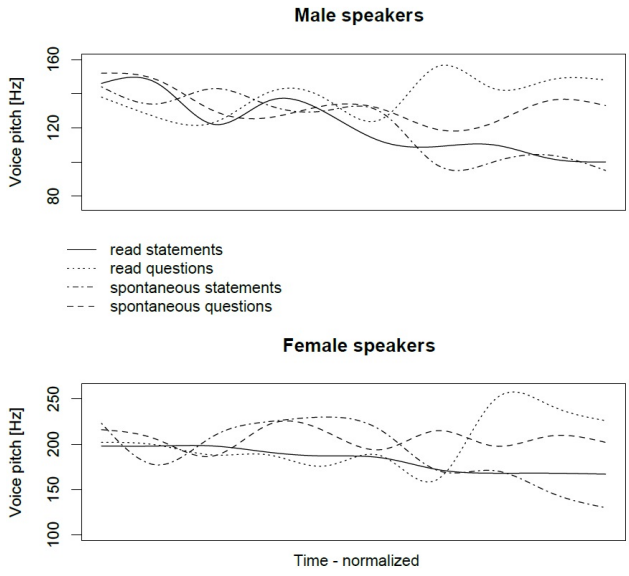
2. MATERIAL

A 45 year-old male and a 30 year-old female speaker recorded the read stimuli as questions and as statements. The stimuli were drawn from [7] and a translation of [9]. There were 11 questions and statements with declarative word order.

The non-scripted questions and statements were drawn from the Danish Phonetically Annotated Spontaneous Speech Corpus [3]. These were 11 declarative phrases that had been consistently identified as questions, and 11 declarative statements [2]. The questions were spoken by 4 female and 4 male speakers, while the 11 statements were spoken by 3 female and 3 male speakers. The speakers ranged in age from 22 to 62 years.

The mean F_0 traces for these stimuli are shown in Fig. 2. When comparing the mean F_0 characteristics of the questions it can be seen that the non-scripted items decline more than the read and thereby follow the observation reported in [2].

Figure 2: Showing the mean F_0 traces for read and non-scripted stimuli used in the utterance type identification, shown over a normalised timescale. Curves derived from [10] and rendered with a spline curve.



For the subject/non-subject identification task, which tested the perception of local pitch movements, six pairs of sentences were recorded. A pronoun was in the initial position of the second sentence of each pair. Manipulation of the prominence of this pronoun varied the antecedent to which it referred, for example:

[Stimulus] Deres sommerhus ligger 70 kilometre fra huset i Gilleleje. $\text{Det}_{[\text{pronoun}]}$ trænger forresten til at blive malet.
Their summerhouse is 70 kilometres from the house in Gilleleje. It $_{[\text{pronoun}]}$ *could do with being painted.*

[Question] Hvilket hus trænger til at blive malet?
Which house could do with being painted?

[Response alternatives] Huset | Sommerhuset
The house | The summerhouse

If the pronoun was perceived as prominent, the non-subject was expected to be the antecedent, i.e., *the house in Gilleleje*. If the pronoun was perceived as not prominent, the subject was expected to be the antecedent, i.e., *the summerhouse*. A 40 year-old male recorded the stimuli with a prominent and not prominent pronoun. The Tandem-STRAIGHT algorithm [6] was then used to morph the F_0, intensity and duration of the pronoun, so as to derive two intermediate prominence levels.

3. METHOD

Listening participants were 22 staff and students from the department. They were aged between 22 and 65 years (mean 34, SD 14) and 19 were female. In a pre-test audiometric screening, all participants were found to have bilateral pure-tone thresholds lower than 25 dB HL at octave frequencies from 250 to 2000 Hz,

The utterance type and subject/non-subject stimuli were presented randomly in three conditions i) unprocessed ii) Eight-channel noiseband vocoded, and iii) Eight-channel vocoded where the lowest band was unprocessed and the higher bands were noiseband vocoded. The low pass cutoff frequency of the lowest band was set to 300 Hz. This cutoff was chosen, as it was the mean plus one SD of the maximum F_0 of all utterance type stimuli. Higher frequency bands were then spaced according to the Greenwood function [1].

3.1 Procedure

The stimuli where presented to participants via headphones in a sound-treated room. A computer monitor displayed the response alternatives and participants responded on a keyboard. During the utterance type identification, participants responded if the item sounded like a question or if it did not. For the subject/non-subject identification

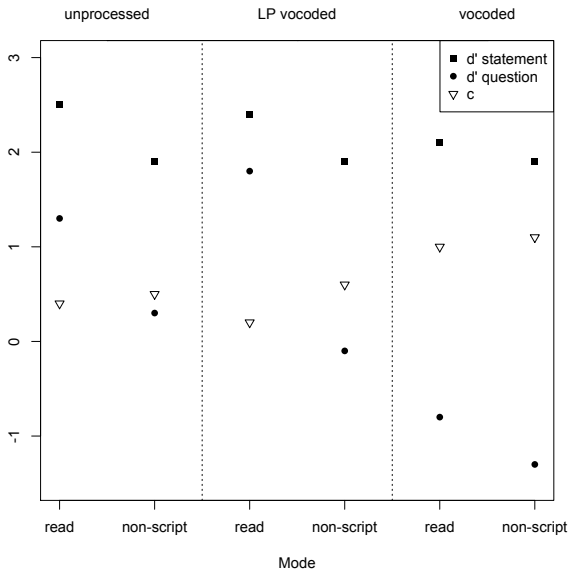
participants selected the leftward antecedent that they judged as having been inferred. The two tasks were counterbalanced across participants and stimuli were presented randomly in each task.

4. RESULTS

4.1 Utterance type identification

Mean correct utterance type identification across the three processing conditions was significantly different for the read and non-scripted stimuli (Table 1). Mean performance across all processing conditions was at 81% (SD 39) for the read material, and at 66% (SD 47) for the non-scripted material. We examined the mean sensitivity (d') to questions and statements (Fig. 3) and observed that sensitivity to non-scripted questions was consistently the lowest for all processing conditions. A mean effect of utterance type was also found (Table 1). To investigate this, the mean response bias (c) was calculated for statements and found to be greatest for the vocoded condition. This indicates that with the degraded speech signal the mean bias of participants increased towards the statement response alternative.

Figure 3: Mean d' and c for question and statement identification according to mode (read/non-scripted) and processing condition.



4.2 Subject/non-subject identification

Mean results showed that subject identification was greater with the not prominent pronoun and non-subject identification was greater with the most prominent pronoun (see Fig. 4). Responses to items with intermediate prominence formed a continuum between the non-prominent and the most prominent endpoints. Total responses to stimuli with prominent and non-prominent pronouns that were below the random response rate (0.57) were omitted. Performance on the low-pass vocoder and the unprocessed conditions was very similar indicating that antecedent signalling was conveyed largely by F_0 variation.

Figure 4: Showing the percentage of subject and non-subject identification for unprocessed (squares), low-pass vocoded (triangles) and vocoded (crosses) stimuli. NB, vocoded stimuli were measured only with items containing naturally prominent and not prominent pronouns.

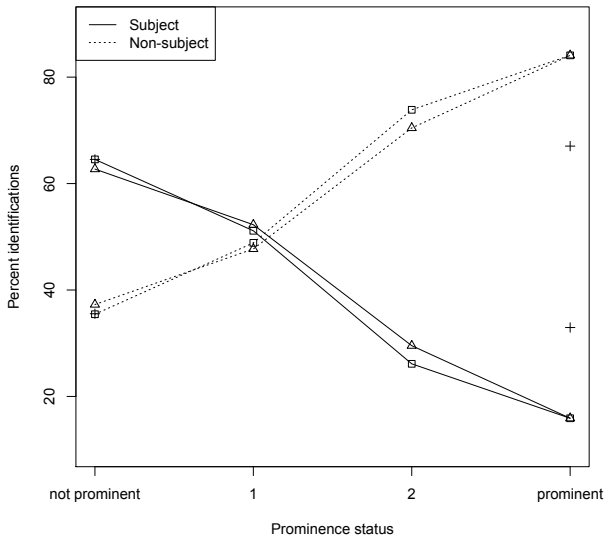


Table 1: Showing the fixed factors from the mixed-logit model (2904 observations, n=22). These are mode of presentation (reference read), utterance type (reference question), and not prominent/non-subject and prominent/subject identification.

Factor	Coef.	S.E.	z	*p*
Mode	-1.1	0.3	-3.9	<0.001
Utterance type	2.9	0.3	10.0	<0.001
Not prominent	-0.7	0.8	-0.9	n.s.
Prominent	0.6	0.8	0.8	n.s.

4.3 Performance comparisons

The utterance type identification responses were submitted to mixed logistic regression modelling with participant and item as crossed random effects (cf. Table 1). The fixed factors included the individual proportions of non-subject identifications of prominent pronouns, and subject identifications of not prominent pronouns. Both of these factors were

not significant, indicating that there was no associative link between performance on these two tasks. The model also revealed significant main effects for mode (non-scripted and read) and statement identification.

To compare performance on the non-scripted and the read question/statement identification we compared the individual d' values of question identification in the unprocessed condition and found that these were not correlated (r=-0.0820, p=0.7).

5. DISCUSSION

Data from this study show a significant difference in utterance type identification between read and non-scripted speech, along with no correlation between the sensitivity of listeners to these two modes. This suggests that testing question and statement identification with read material does not provide results that reflect the ability of a listener in everyday communication. The measure of group bias towards the non-question like response alternative was higher for the non-scripted than the read material. Also, *c* values were similar between the unprocessed and low-pass vocoded conditions and increased when the intonation contour was marred in the full-vocoded condition. This shows that in the absence of a distinct intonation contour, participants were more likely to identify stimuli as statements and this finding is in line with previous work that implicates statements as the unmarked response alternative [5, 8].

Our hypothesis regarding a link between performance on utterance type and subject/non-subject identification was not borne out and we report no link between these two measures. This may be due to two, possibly inter-related, reasons. Firstly, the perceptual task involved in utterance type and subject/non-subject identification may tap essentially different underlying language abilities. That is, utterance type identification requires a listener to attend to the global intonation contour in order to identify the utterance type, while the semantic content of the material can be largely ignored. Conversely, subject/non-subject identification requires the listener to infer an antecedent from the level of prominence, and thus the content of the material and the language diligence of the listener may play a role in this task. Secondly, the perception of global intonation contours that convey utterance type seems to be unrelated to the perception of local prosodic variation that occurs over a shorter time-course. This reason could support the view that there are implicit perceptual distinctions between the components of the hierarchical superpositional model of Danish prosody [4].

6. REFERENCES

[1] Greenwood, D. D. 1961. Critical bandwidth and the frequency coordinates of the basilar membrane. *J. Acoust. Soc. Am.* 33, 1344–1356.

[2] Grønnum, N., Tøndering, J. 2007. Question intonation in non-scripted Danish dialogues. *Proc. 16th ICPhS* Saarbrücken, 1229–1232.

[3] Grønnum, N. 2015. DanPASS – Danish Phonetically Annotated Spontaneous Speech. www.danpass.dk.

[4] Grønnum, N. 1995. Superposition and subordination in intonation – a non-linear approach. *Proc. 13th ICPhS* Stockholm, 2, 124-131.

[5] Grønnum, N. 1992. The groundworks of Danish intonation. Museum Tusculanum: Copenhagen.

[6] Kawahara, H., Takahashi, T., Morise, M., Banno, H. 2009. Development of exploratory research tools based on TANDEM-STRAIGHT. *Proc. APSIPA ASC,* 111-120.

[7] Lieberman, S., Michaels, T. 1962. Some aspects of fundamental frequency and envelope amplitude as related to the emotional content of speech. *J. Acoust. Soc. Am.* 34, 922-927.

[8] Ma, J. L.-Y., Ciocca, V., Whitehall, T. 2011. The perception of intonation questions and statements in Cantonese. *J. Acoust. Soc. Am.* 129, 1012-1023.

[9] Møller, V., Post, I. 1997. Voksne med cochleaimplantat: materiale til iagttagelse [Adults with cochlear implants: material for observation].

[10] Xu, Y. 2013. ProsodyPro-A tool for large-scale systematic prosody analysis. *Proc. TRASP* Aix-en-Provence, France 7-10.